Application/Control Number: 09/996,081
Art Unit: 2655

Docket No.: 2000-0573

REMARKS

Reconsideration and allowance are respectfully requested. Claims 1 – 16 are pending and claims 1, 2, 8 and 14 are amended.

Rejection of Claims 1, 8 and 12 – 14 Under Section 102

The Examiner rejects claims 1, 8 and 8-14 under Section 102 as being unpatentable over U.S. Patent No. 5,864,810 to Digalakis et al. ("Digalakis et al."). Applicants traverse this rejection and submit that Digalakis et al. fail to teach each claim limitation. Although Applicants do not agree with the Examiner's analysis of Digalakis et al., Applicants have amended claims 1, 8 and 14 without prejudice or disclaimer in order to further prosecution of the case. Claim 2 is amended to promulgate the changes made to claim 1. Support for the amendment is found at least in paragraph [0075] of the application.

Amended claims 1, 8 and 14 recite generating a plurality of lattices for received speech utterances associated with filling in a plurality of data fields. In no place do Digalakis et al. discuss the process of performing speech recognition in the context of filling in a plurality of data fields. Form filling differs in terms of speech recognition in contrast to recognition in the context of a natural language dialog. As previously discussed, Digalakis et al. focus on a technique for performing ASR that adapts to a particular speaker by using adaptation data to develop a transformation through which speaker independent models are transformed into speaker adapted models. They then teach about how the speaker adapted models are used for speaker recognition. Inasmuch as the focus of Digalakis et al. is primarily on speaker identification and improved ASR for that identified speaker, Applicants submit that a review of this reference reveals that they do not focus on ASR in the context of filling in a plurality of data fields. Therefore, they fail to teach the steps (with reference to claim 1) of generating a plurality of lattices for received speech utterances associated with

filling in a plurality of data fields and then concatenating the plurality of lattices into a single concatenated lattice.

The Examiner had pointed to col. 11, lines 40-44 of Digalakis et al. as teaching generating lattices for speech utterances. There, they teach:

> For fast experimentation, we used the progressive search framework: an initial, speaker-independent recognizer with a bigram language model outputs word lattices for all the utterances in the test set. These word lattices are then rescored using speaker-dependent or speaker-adapted models. We performed two series of experiments, on native and non-native speakers of American English, respectively. All experiments were performed on the 5,000-word closed-vocabulary task, and are described below.
>
> To compare SI, SD and SA recognition performance on native speakers, we performed an initial study of our adaptation algorithms on the phase-0 WSJ corpus.

They then describe how the systems were trained using 3,500 sentences from 42 male speakers and so forth. It is clear that the word lattices and for utterances in the test set (the Wall Street Journal corpus was used) are drawn from a large corpus and are provided for general speech recognition rather than the different context of form filling. One of skill in the art understands that the ASR models and speech recognition process will differ given these different purposes.

Furthermore, the last step of claim 1 requires applying at least one language model to the single concatenated lattice in order to determine a relationship between the plurality of lattices. Inasmuch as Digalakis et al. fail to teach anything regarding ASR for filling in data fields, the also fail to teach this steps of determining a relationship between a plurality of lattices as is recited in claim 1.

Regarding the steps of concatenating the lattices, we note that the Examiner pointed to col. 6, lines 45 - 53 and col. 12, lines 56-59. In column 6, Digalakis et al. teach creating a set of "tied" models from the trained models 117. The trained "SI" models 117 are the "speaker independent models" that a previously created and stored when doing speech recognition. See col. 3, lines 13 - 41 wherein they discuss how the SI training data are speech samples from a database of samples from different speakers. Note the contrast to the "SD" (speaker

dependent) data associated with speech recognition of a speaker. These are previously stored and not generated lattices for speech utterances associated with filling in a plurality of data fields. Regarding the citation to column 12, the concept of "clustering" in Digalakis et al. is for the purpose of selecting "an arbitrary degree of mixture tying across different HMM states". Col. 12, lines 47 - 49. This concept differs from the step of claim 1 which requires concatenating the plurality of lattices into a single concatenated lattice, in which case the plurality of lattices are generated for received utterances associated with filling in a plurality of data fields. Digalakis et al. is analyzing and focusing on speaker adaptation for native and non-native speakers. Therefore, the steps of claim 1 are not taught by Digalakis et al. and claims 1, 8 and 14 are patentable and in condition for allowance.

Claims 12 and 13 depend from claim 8 and recite further limitations therefrom. Therefore, these claims are patentable as well over Digalakis et al.

Rejection of Claims 1, 8 and 12-14 Under Section 102

The Examiner alternatively rejects claims 1, 8 and 12 - 14 under Section 102 as being anticipated by U.S. Pat. No. 5,848,389 to Asano et al. ("Asano et al."). Applicants submit that the amended claims are patentable over this reference.

For example, in claim 1, the Examiner points to column 5, lines 21 - 37 as teaching generating lattices. However, Applicants submit that Asano et al. fail to teach generating a plurality of lattices for received speech utterances associated with filling in a plurality of data fields. With regards to the concatenation step of claim 1, the Examiner points to column 7, lines 1 - 12. This portion of Asano et al. states:

> The proposed word subject lattice outputted from the recognizing unit 4 is supplied to the example retrieving unit 5. Upon receipt of the proposed word subject lattice, the example retrieving unit 5 performs a process operation in accordance with the flow chart of FIG. 3, for example. That is to say, first, at step S1, the words for constituting the word lattice are combined with each other, and then a word column or series (sentence) made of at least one word is formed. It should be noted that at this time, the

words of the proposed word subject lattice do not overlap each other along the time based direction, and that they also are combined with each other in the time sequence.

It is clear from this citation that Asano et al. fail to teach concatenating lattices together. Rather, Asano et al. teach words that are used to "constitute" the word lattice are combined with each other. The Examiner appears to be confusing concatenating two lattices together with the process discussed in Asano et al. where they combine words together to create a single word lattice. Furthermore, this portion of Asano et al. clearly fails to teach concatenating a plurality of lattices into a single concatenated lattice wherein each lattice of the plurality of lattices are associated with utterances received for filling in a plurality of data fields.

Finally, the Examiner asserts that column 11, lines 7 - 22 teach the last step of claim 1. This portion of Asano et al. fail to teach applying at least one language model to the single concatenated lattice in order to determine relationships between the plurality of lattices. As mentioned above, Asano et al. fail to teach concatenating more than one lattice together (they teach combining words to create a word lattice). Therefore, they fail to teach applying at least one language model to the concatenated lattice as is recited in the claims.

Applicants respectfully submit that Asano et al. fail to teach claim 1, either previous to the present amendment, and/or especially in view of the present amendment to this claim. Similarly, this approach applies to claims 8 and 14. Therefore, claims 1, 8 and 14, including claims 12 and 13, are patentable and in condition for allowance in view of Asano et al.

Rejection of Claims 1, 8 and 13-14 Under Section 102

The Examiner also rejects claims 1, 8 and 13 - 14 as being anticipated by U.S. Patent No. 6,581,033 to Reynar et al. ("Reynar et al."). Applicants submit that Reynar et al. fail to teach each limitation of the claims.

We first turn to claim 1. The Examiner asserts that Reynar et al. teach in col. 8, lines 43 - 46 teach the step of generating lattices. Applicants note that this portion of Reynar et al. teach storing nodes within a lattice where the nodes are associated with a probability assigned to the associated word or phrase. The lattice is then traversed to produce likely alternatives for any section of text. They mention that adjacent pieces of text can be combined into a larger lattice through a process known as concatenation. What Reynar et al. fail to teach is generating a plurality of lattices wherein each lattice is from received speech utterances associated with filling in a plurality of data fields. Reynar et al. do not teach or suggest the data field input context of claim 1. Notably, Reynar et al. teach away from such an application in that their whole approach is focused on correcting speech recognition *mode* errors. Modes include a "command" mode or a text entry mode. See col. 8 in general where they teach how to make corrections if speech is received in a command mode (such as 'delete the previous word' as a command to do that action) and a text entry mode (such as 'delete the previous word' to be added as text to a document). If the system erroneously deleted the previous word instead of inserted the text "delete the previous word" into the document, then a mode correction processor would undo the erroneous command (by undeleted the previous word) and then take the proper action (by inserting the text into the document).

Where Reynar et al. focus on how to switch and make mode corrections as described above, Applicants submit that they do not teach or suggest the first step of claim 1.

Similarly, in column 8, Reynar et al. fail to teach concatenating the plurality of lattices into a single concatenated lattice - taking into account that the plurality of lattices recited in this claim set are lattices associated with speech for data fields. As discussed above, Reynar et al. teach managing a mode correction processor between a command mode and a text input mode.

Finally, because Reynar et al. fail to teach the first limitation of claim 1 related to the plurality of lattices, Reynar et al. easily fail to teach applying at least one language model to

the single concatenated lattice in order to determine relationships between the plurality of lattices. There is no discussion of any analysis of determining a relationship between a plurality of lattices (each lattice related to filling a data field). Therefore, this step of claim 1 is also not taught by Reynar et al.

For the foregoing reasons, Applicants submit that claim 1 is patentable over Reynar et al. Similarly, claims 8 and 14, as well as claim 13, are also patentable and in condition for allowance.

Rejection of Claims 2, 6 - 7, 9, 11 and 16 Under Section 103

The Examiner rejects claims 2, 6-7, 9, 11 and 16 under Section 103 as being obvious in view of U.S. Patent No. 6,581,033 to Reynar et al. ("Reynar et al.") and U.S. Publication No. 2002/005272 to Thrasher et al. ("Thrasher et al."). Applicants note that the Examiner previously rejected claims 2 - 7, 9 - 11 and 15 - 16 as being unpatentable under Section 103 in view of Digalakis et al. and Thrasher et al. and now uses Reynar et al. in place of Digalakis et al. Applicants traverse this rejection and submit that these claims are patentable for the following reasons.

In the Office Action, the Examiner responded to Applicants numerous arguments regarding why there is a lack of motivation or suggestion to combine Digalakis et al. with Thrasher et al. (as applied to claims 2 - 7, 9 - 11 and 15 - 16) by only noting that the motivation for the combination is found in the previously-cited column and line number. However, in the current office action, no claims are rejected by the combination of Digalakis et al. and Thrasher et al. and therefore Applicants assume at least substantively the Examiner has conceded that there is no motivation to combine these references inasmuch at the Examiner no longer combines these references to reject any claims.

Inasmuch as the features of the parent claims are not taught by Reynar et al. or the other references as discussed above, each of these dependent claims is patentable and in

condition for allowance. In addition, we will address the motivation to combine these references.

We now turn to the combination of Reynar et al. and Thrasher et al. to reject claims 2, 6 - 7, 9, 11 and 16. Applicants note that the standard by which the Examiner ascertains whether motivation to combine exists is only by a "preponderance of the evidence." If the examiner determines there is factual support for rejecting the claimed invention under 35 U.S.C. 103, the examiner must then consider any evidence supporting the patentability of the claimed invention, such as any evidence in the specification or any other evidence submitted by the applicant. The ultimate determination of patentability is based on the entire record, by a preponderance of evidence, with due consideration to the persuasiveness of any arguments and any secondary evidence. *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). The legal standard of "a preponderance of evidence" requires the evidence to be more convincing than the evidence which is offered in opposition to it. With regard to rejections under 35 U.S.C. 103, the examiner must provide evidence which as a whole shows that the legal determination sought to be proved (i.e., the reference teachings establish a *prima facie* case of obviousness) is more probable than not. MPEP 2142.

The test for obviousness is what the combined teachings of the references would have suggested to one of ordinary skill in the art, and all teachings in the prior art must be considered to the extent that they are in analogous arts. Where the teachings of two or more prior art references conflict, the examiner must weigh the power of each reference to suggest solutions to one of ordinary skill in the art, considering the degree to which one reference might accurately discredit another. *In re Young*, 927 F.2d 588, 18 USPQ2d 1089 (Fed. Cir. 1991). MPEP 2143.01.

With these principles in mind, we turn to the analysis of any motivation to combine Reynar et al. with Thrasher et al. Applicants respectfully submit that merely citing, column 3, paragraph 35 in Thrasher et al., is not sufficient to establish the requisite motivation to

combine by a preponderance of the evidence. While this paragraph does mention a "confidence measure", the Examiner states that it would be obvious to use such as confidence score for "identifying which patterns are most likely to have been improperly identified by the recognizer." However, the ability to identify which patterns are most likely to have been properly identified is already accomplished in Reynar et al. In column 8, lines 41-54 for example, Reynar et al. state that each word or phrase of speech input data is assigned a probability or placed in an "n-best" alternatives list that includes the n-best alternatives to each word or phrase. Reynar et al. already has the ability within the scope of its teachings to "identify which patterns are most likely to have been properly identified by the recognizer." This is accomplished by the n-best list. The first recognized word or phrase is the most likely correct recognition based on the probability analysis. The second recognized word or phrase is the second-most-likely to be the correctly recognized word or phrase, and so forth.

In this regard, the very motivation to combine articulated by the Examiner, i.e., the reason, feature or benefit that the Examiner states should be brought into Reynar et al. from Thrasher et al. is already found in Reynar et al. Therefore, incorporating the confidence score of Thrasher et al. would be duplicative of the n-best alternatives list produced by Reynar et al. When weighing the motivation value then of these two teachings, Applicants submit that the preponderance of the evidence would lead away from concluding that one of skill in the art would combine these references.

Applicants incorporate the previous arguments presented wherein we argued that the overall suggestive power of each reference with regards to the focus and primary teaching of each reference do not lead to a suggestion to combine. In sum, When the entire teachings of the prior art are considered for their suggestive power with regards to combining with each other, they do not suggestion or provide motivation to blend their teachings.

Reynar et al. have as their main focus the ability to make recognition mode error corrections. The example discussed above between a "command" mode and a text insertion mode is sufficient to make clear the focus of Reynar et al. In contrast, Thrasher et al. focus on a method of generating alternatives to words indicative of recognized speech. They address the problem of correction of recognized text and require an operator selection of input to identify a selected portion of recognized speech for which alternatives are generated. Paragraph [007]. Where the ASR module has provided text from the speech, and the user identifies various words or phrases in the text for correction, Thrasher et al. teach that during this correction process, the system presents alternative words for recognition for each word the user desires to correct. For large dictations, it becomes cumbersome to maintain all the possible alternatives for each possible word to be corrected. Therefore, their invention focuses on a method of generating alternatives to words indicative of recognized speech to help in the text correction of speech. Necessarily, as part of their invention, the user must select a portion of the recognized speech for correction. Paragraphs 0005 - 0007.

The fact that Thrasher et al. teach that a user will select text for correction is instructive and teaches away from blending its teachings with Reynar et al. When in the "command" mode of Reynar et al., there is no presentation of text to the speaker. In the command mode, where a user says "delete the previous word" there is no desire on the user's part to see those words appear in a document. Rather, the user desires an action to be taken in obedience to that command. When in the dictation mode, Reynar et al. already teach using a correction method in which the user is presented a candidate selection as a highlighted choice and other alternative results are presented. The user can then select the candidate or another alternative in the results to select the desired results. See Col. 11, lines 46 - 61.

The Examiner asserts that a person of skill in the art would find it obvious to modify Reynar's method to generate a confidence score to identify which patterns are most likely to have been improperly identified by the recognizer. However, this ability is already taught in

Reynar et al. with its candidate selection and alternative dictation results listing available to the user for selection. Furthermore, given this ability already taught by Reynar et al., one of skill in the art would not likely (under the preponderance of the evidence) be motivated to add Thrasher et al.'s generation of alternative words for user selection inasmuch as Reynar et al. already has that feature. Given the overall suggestive power of each reference, Applicants respectfully submit that there is insufficient motivation or suggestion in these references to combine them.

Therefore, claims 2, 6 - 7, 9, 11 and 16 are patentable and in condition for allowance.

Rejection of Claims 3 - 4 Under Section 103

The Examiner rejects claims 3 and 4 under Section 103 in view of Reynar et al., Thrasher and Waibel et al. Applicants incorporate the above arguments and submit that (1) because the parent claim is patentable and (2) because there is insufficient motivation to combine Reynar et al. with Thrasher et al, that these two claims are patentable and in condition for allowance.

Rejection of Claim 5 Under Section 103

The Examiner rejects claim 5 under section 103 as being unpatentable under Reynar et al. with either Morin et al, Flanagan et al. or L'Esperance. Claim 5 adds the limitation of rescoring after a speech recognition model has been compensated to reflect acoustic environment data and transducer data. Applicants respectfully submit that inasmuch as the parent claim 1 is now patentable and in condition for allowance per the discussion above, that this dependent claim 5 is also patentable. Applicants do not concede, however, that it is obvious or appropriate to combine any of these references.

Application/Control Number: 09/996,081
Art Unit: 2655

Docket No.: 2000-0573

CONCLUSION

Having addressed the rejection of claims 1 - 16, Applicants respectfully submit that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited.

Respectfully submitted,

Date: September 9, 2005

By: /Thomas M. Isaacson/

Correspondence Address:
Samuel H. Dworetsky
AT&T Corp.
Room 2A-207
One AT&T Way
Bedminster, NJ 07921

Thomas M. Isaacson
Attorney for Applicants
Reg. No. 44,166
Phone: 410-414-3056
Fax No.: 410-510-1433